PROCEEDINGS OF TDWG

Biodiversity Information Science and Standarts 1: e20608
doi: 10.3897/tdwgproceedings.1.20608

OPEN ACCESS

Conference Abstract

SeqDB: Biological Collection Management with Integrated DNA Sequence Tracking

Satpal Bilkhu‡, Nazir El-Kayssi‡, Matthew Poff‡, Anthony Bushara‡, Michael Oh‡, Joseph Giustizia‡, Iyad Kandalaft‡, Christine Lowe‡, Oksana Korol‡, Joel Sachs‡, Keith Glen Newton‡, James Macklin‡

‡ Agriculture and Agri-Food Canada, Ottawa, Canada

Corresponding author: Satpal Bilkhu (satpal.bilkhu@agr.gc.ca)

Received: 25 Aug 2017 | Published: 26 Aug 2017

Citation: Bilkhu S, El-Kayssi N, Poff M, Bushara A, Oh M, Giustizia J, Kandalaft I, Lowe C, Korol O, Sachs J, Newton K, Macklin J (2017) SeqDB: Biological Collection Management with Integrated DNA Sequence Tracking . Proceedings of TDWG 1: e20608. https://doi.org/10.3897/tdwgproceedings.1.20608

Abstract

Agriculture and Agri-Food Canada (AAFC) is home to a world-class taxonomy program based on Canada's national agricultural collections for Botany, Mycology and Entomology. These collections contain valuable resources, such as type specimen for authoritative identification using approaches that include phenotyping, DNA barcoding, and whole genome sequencing. These authoritative references allow for accurate identification of the taxonomic biodiversity found in environmental samples in fields such as metagenomics. AAFC's internally developed web application, termed SeqDB, tracks the complete workflow and provenance chain from source specimen information through DNA extractions, PCR reactions, and sequencing leading to binary DNA sequence files. In the context of Next Generation Sequencing (NGS) of environmental samples, SeqDB tracks sampling metadata, DNA extractions, and library preparation workflow leading to demultiplexed sequence files. SeqDB implements the Taxonomic Databases Working Group (TDWG) Darwin Core standard Wieczorek et al. 2012 for Biodiversity Occurrence Data, as well as the Genome Standards Consortium (GSC) Minimum Information about any (X) Sequences (MIxS) specification Yilmaz et al. 2011. When coupled with the built-in data standards validation system, this has led to the ability to search consistent metadata across multiple studies. Furthermore, the application enables tracking the physical storage of the

© Bilkhu S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

aforementioned specimens and their derivative molecular extracts using an integrated barcode printing and reading system.

All the information is presented using a graphical user interface that features intuitive molecular workflows as well as a RESTful API that facilitates integration with external applications and programmatic access of the data. The success of SeqDB has been due to the close collaboration with scientists and technicians undertaking molecular research involving the national collection, and the centralization of their data sets in an access controlled relational database implementing internationally recognized standards. We will describe the overall system, and some of our lessons learned in building it.

Keywords

Database, Bioinformatics, Sanger, Next Generation Sequencing, Biodiversity, Metagenomics, DNA Sequencing

Presenting author

Satpal Bilkhu

References

- Wieczorek J, Bloom D, Guralnick R, Blum S, Doering M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: an evolving community-developed biodiversity data standard. PLoS One 7: 1-8.
- Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, G Chain PS, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, Gil IS, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spor A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29 (5): 415-420. https://doi.org/10.1038/nbt.1823